DOCUMENT RESUME

ED 179 729 CE 023 250

AUTHOR	Warren, Sue Allen; Gardner, David C.
TITLE	Vocational Training for Disabled Students. A Handbook on Assessment.
INSTITUTION	Blue Hills Regional Technical School, Canton, Mass.; Boston Univ., Mass. Dept. of Business and Career Education.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C.
BUREAU NO	498AH50296
PUB DATE	Jun 77
GRANT	G007500558
NOTE	66p.; For related documents see CE 023 251-257
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Aptitude Tests; Experience; Guidelines; *Handicapped Students; Interest Tests; *Predictive Measurement; Secondary Education; *Success Factors; *Testing; *Test Selection; Test Validity; *Vocational Education

ABSTRACT

The Vocational Instruction, Training and Assessment (VITA) Project conducted a study to determine whether a battery of tests used with a sample of handicapped students enrolled in vocational education program(s) could aid in predicting success in vocational training. Procedures, instruments, and student samples were selected to provide data on (1) interest measures; (2) aptitude measures; and (3) short-term experiences within specific training programs. The sample consisted of ninth- and tenth-grade students from two different vocational-technical high schools. Among the findings are the following: (1) a large number of published tests seem to have potential usefulness with one or more of the handicapped populations (list provided); (2) the two most powerful predictors of class rank were IQ and handicapping condition; and (3) selection of measurement devices must be directly related to the questions one has (e.g., Can the student learn the skill?, What interests does the student show that have implications for choice of training programs?). (This study is presented in the format of a handbook which also provides information on the purposes, utilization, and selection of vocational measurement instruments.) (BM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED179729

VOCATIONAL TRAINING FOR DISABLED STUDENTS

A Handbook on Assessment

Sue Allen Warren and David C. Gardner
Departments of Special Education and Business and Career Education
School of Education
Boston University
765 Commonwealth Avenue
Boston, Massachusetts 02215

Prepared for:

Project VITA: A Cooperative Activities Model for University and Vocational/ Technical School Collaboration, USOE Project No. 498AH50296, Grant No. G007500558.

CE 023 250

Published by: Department of Business and Career Education . School of Education . Boston University . Boston, Massachusetts

JUNE 1977

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

TABLE OF CONTENTS

		Page
I.	Purposes of Assessment	1
II.	Measures Commonly Used in General Practice	4
	Validity of Measures	6
	Using Tests	8
	Selection of Measurement Tools	10
III.	Selected Vocational Measures	13
	Introduction	13
	A. Specific Occupations	14
	B. Aptitude Batteries	16
	C. Mechanical Ability	19
	D. Manual Dexterity	21
	E. Interests	23
	F. Miscellaneous	31
IV.	Tests Considered for Project VITA	34
V.	Methodology and Results	39
	Introduction	39
	Measures Selected	41
	Method	44
	Results of Analysis of Data	45
	Prior School Grades and Class Rankings Assigned	48
	Interpretation	48
	Intelligence Test Results	49
	Overview	57
VI.	Summary and Recommendations	60
	References	62

I. PURPOSES OF ASSESSMENT

Assessment has numerous purposes. It is valuable for program evaluation, individual evaluation, and for prediction.

As a part of Project VITA, assessment of the modules designed for the study was done. The general procedure has been described in the final report (Gardner & Kurtz, 1977). Because of the "pilot" nature of this project and the fact that it was done in public schools where assignment to experimental and control groups was not feasible, it was not possible to use an experimental design that gave comparative data. Thus, the program evaluation suffers from lack of precision in both experimental design and in measurement devices used. One can report, however, that using a criterion-reference measure to determine whether these students mastered the materials presented, it was successful. The criterion for selection for inclusion of a disabled student in the project was difficulty in learning the materials being presented in the graphic arts shop. It may be that these students might have learned the materials eventually without the modules. Perhaps they would have learned it as rapidly, but on the basis of teacher reports, this appears to be unlikely. Furthermore, the ratings by teachers and students indicated that the students did enjoy the materials and did rate the modules higher than other shop instruction. In general, the students were highly positive in their ratings, students did learn the materials, even though they were chosen because they were having difficulties with these concepts. Thus, within the limitations discussed above, one can call the program a success.

Individual evaluations of students is another major use of assessment. Within the limits of this project, these assessments were primarily efforts to

determine the relationships between various published measures and student performance in shop classes, as determined by teacher ranking of position of students in classes and by teacher ratings of student work performance. These findings are reported in some detail in Section V of this report. The crucial test, later performance in the world of work, was not included in this study. To determine job performance requires a longitudinal study, not a brief one of the scope possible in the time limits of this project. However, a continuation of this work, using students from one of the schools participating in this project, is now underway and may offer some indication of the degree to which such factors as teacher assigned grades and class rank are predictive of job performance in the community after graduation.

All measurement devices have limitations. Humans are highly complex and highly individual. No single measure or set of measures can be expected to provide precision in educational planning. It is possible, however, to determine certain procedures that can reduce the "best guess" or "clinical evaluation" that is often used in planning. One can determine "test error" or the degree of latitude one must consider in using tests. One can determine cut-off points below which success is highly unlikely. One can gain some understanding of relationships between work performance and other variables. Such determinations are highly useful, though admittedly imperfect. The danger in using published measurement devices lies less in the measures themselves than in the ways in which such measures are used by planners who feel strong needs to improve their current predictions or their current procedures. If one can recognize the probability nature of all assessment, whether this be clinical observations or standardized measures, then one can use assessment procedures with profit. Those who seek total accuracy or extreme precision will be disappointed. The proof of any assessment procedure is the degree to which it is useful. In the

procedures used for this project, some measures were found to be useful, others were found not to be very useful, and still others were equivocal (see Section V).

Assessment procedures that are useful for some purposes are not necessarily useful for others. Furthermore, the title of a test published as a measure of a particular trait or variable does not necessarily indicate that the test is reliable or valid for the purposes for which it is selected by an examiner (or even the purposes for which it is said to be intended). In this, as in other aspects of life, the rule must be "caveat emptor;" let the buyer beware. The proof of a measure's usefulness must be demonstrated in individual situations. Some measures, such as the Stanford-Binet Intelligence Scales and the Wechsler Scales of Intelligence, have been subjected to thousands of studies to determine their usefulness. The same cannot be said of many of the many thousands of tests available on the commercial market and those developed for specific research purposes.

In this project, an effort was made to find measures that might be related to success in vocational training of handicapped students. Generalizations from these findings must be considered in the perspective of similarity of other populations to these students. It seems reasonable to suppose that the vocational students who participated in this project may be somewhat similar to other handicapped students in vocational training programs in some other schools in America, but the degree of generalizability must be considered in terms of the degree of similarity between schools, students, programs, and teachers.

Despite the limitations of published measures, many of them are probably quite helpful when used with due regard for the problems of all observations, including direct observations.

II. MEASURES COMMONLY USED IN GENERAL PRACTICE

One of the problems in selection of assessment measures is the plethora of them. Two tests with highly similar names and presumably intended to measure similar variables may not show strong correlations with each other when subjected to tests. For example, many of the measures (most of them developed for experimental purposes) that bear some form of the term "self-concept" in the title show low correlations with each other when both are presented to the same population. As discussed later in this report, two measures that are designed to measure some aspects of interests were not highly related to each other for vocational students in two schools in Massachusetts.

Consistency

Perhaps the most consistent measures, at least as indicated by correlation studies, are the general intelligence scales which have been under development for decades. For the Stanford-Binet Intelligence Scale, many studies report very high reliabilities (with various procedures), correlations ranging from the high eighties to the low nineties. The same can be said for the various Wechsler Scales of Intelligence. Correlations between these two scales are also quite high. Standardized measures of academic achievement also tend to show good test-retest reliabilities and relatively good consistency across the more comprehensive measures when they are compared with each other. High reliability, however, does not ensure similar findings; one of the more clear examples of this is the repeated reports of fairly high correlations between the Binet or the Wechsler Scales and the Peabody Picture Vocabulary Test (PPVT); however, many other studies have indicated a consistent trend for the Peabody to provide higher IQ scores than either of the other two scales. In its initial development, the PPVT yielded a "Vocabulary Quotient" rather than an "Intelligence

Quotient;" if that term had been retained, perhaps studies of comparison would have been done between measures of vocabulary recognition and the PPVT rather than full scale IQ and the PPVT; such studies may have given more consistency.

Another problem in consistency is demonstrated by studies of test-retest administrations. A group of individuals will be asked to take a test once and then will be retested at some later time (e.g., after one day, one week, one month). The procedure is intended to provide information on the degree to which a test is consistent across administrations, the degree to which some change in score is perhaps due to chance factors. The correlations obtained when this procedure is used vary considerably and are usually reported as "test-retest reliability." If the particular day or week in which a test is taken has a marked influence on the score obtained, one must be particularly cautious in its use. However, the initial test experience may influence later scores of some individuals tested more than other students, leading to lowered test-retest correlations; for example, highly intelligent or curious individuals taking an achievement test might be more inclined to try to determine the "correct" answers to some remembered items that were failed than would less curious or intelligent individuals, leading to greater changes in scores for some individuals than for others.

Another procedure for determining the consistency of a test is to administer it only once, then divide it into two parts for purposes of determining consistency (e.g., split-half reliability). For example, scores might be summed for all even-numbered items and for all odd-numbered items, yielding two scores for each individual. These two "sub-scores" could then be analyzed to determine the degree to which the two halves of the test yield similar scores. When this procedure is used, the experimenter usually employs a "correction factor" to correct for the briefer length of the two halves of the test (as compared to the total length of the full test).

Various other procedures have been developed to assess consistency of tests. The most commonly used, however, are the test-retest and the split-half procedures.

Validity of Measures

In addition to concerns with the degree to which tests are reliable, those who use them are concerned with the validity of such measurement devices. Validity refers generally to the extent to which the test measures "what it is supposed to measure," as contrasted to the question of reliability...whether the test measures consistently. Although it is possible, it is rare that a test can have validity without reliability. Those who construct tests usually find it easier to demonstrate reliability than validity. However, without some evidence that a test does in fact measure what it is intended to measure, it may be of extremely limited usefulness.

In planning for development of a testing device, the constructor of the test is likely to begin with at least some "face validity." That is, on the face of it, the test appears to be a reasonable approach to measuring the variable under consideration. One would not, for example, plan on using a test in which individuals were asked to solve arithmetic problems if the primary task is to determine art appreciation; even if there is a relationship between math skills and art appreciation, this would be a roundabout way of measuring. One begins to develop a test of art appreciation by working at something related more directly to art. Some empirically derived tests may be partial exceptions to this general rule. Occasionally, one may find with serendipity that a test does measure something not anticipated. (More often, one is likely to find that a test does not measure what it was expected to measure.)

Sometimes, a somewhat indirect approach may be used. For example, some of the sub-scales of the Holland VPI were developed on the premise that interests

reflect personality and so certain scales are said to reflect personality variables even though the test is labeled a vocational preference measure. At times, indirect measures have been used to avoid resistance to taking the test on the part of the individual being asked to take it; this approach was, in the past, one used fairly commonly when one was interested in certain personality variables. (Since the emphasis on civil rights and the attack on the testing movement, such procedures have become increasingly rare.)

At other times, a brief measure will be used when one wishes to save the time required to use a long evaluation procedure. For example, if one were interested in determining the potential success of an individual (predictive validity) for airplane pilots, he might develop a long procedure that involved measurement of navigational knowledge, or spatial relationships, and of geographical features. However, if it could be determined that no individuals above a certain level on a shorter and simpler measure had been known to succeed in pilot training programs, then that shorter procedure could be used for a screening procedure. This would not ensure that an individual would succeed, but it could eliminate failures. Such procedures were, in fact, developed during World War II for screening trainees for pilot training. For predictive validity, one is concerned with what one needs to know in order to predict, not what specific types of performance are required by the test. If a test that required one to stand on his head could predict success in algebra, then it would be valid for the purposes, even though common sense suggests that it is a silly approach. The proof is in the predictability.

One rather popular method of attempting to determine if a new test is valid is to compare its results with those of another test that has proven usefulness. In the early days of constructing group intelligence tests and "short" intelligence tests, it was common practice to give the same group of individuals

both the new test and the well established Stanford-Binet Intelligence Scale. If the scores obtained were highly correlated, then one had some feeling of confidence in the new measure. The Binet scale was validated in early studies by comparison of obtained scores with academic achievement; results indicated that it was a good predictor of the academic achievement pupils would make... those with low IQ made less progress academically during a period of instruction than did those with high IQ and the higher the IQ, the greater the progress. However, predictions were imperfect; some individuals with IQ of 95 made greater progress than others with IQ of 106. Few cases were found in which those with 95 surpassed those with 126, however. Because the Binet seemed to predict academic achievement relatively well, tests that were highly correlated with it were assumed to also predict academic achievement; another way of determining the validity of the new test would be to determine directly how well it predicted academic achievement.

Some tests are evaluated on the basis of the degree to which they predict actual behavior that is observed. Scores on a test designed to measure mechanical aptitude could be compared with observations of groups of individuals who are manipulating mechanical apparatus; direct observations of aggressive acts could be compared with scores on a written test intended to measure aggression; sociometric scales in which individuals state their preferences for others in the group can be compared with direct observations of the popularity and exclusion of those individuals. If those who are given "star" ratings on the sociometric scale are also observably the most popular in the group and those who are "isolates" are never included, then one can have some confidence in the validity of the sociometric measure.

Using Tests

Human behaviors are highly complex and complete accuracy is too much

to ask of a test. One value of tests lies in their ability to give information needed in a simpler or briefer or easier way than can be done by longer direct observations. It is possible to determine that a pupil is failing in first grade after he has been in there for a few weeks or months; if one can predict such potential failure with a test requiring one hour (and a highly experienced examiner), then the test may be well worth its costs in time and money to both the educators and the pupil, particularly if one knows some measure to help avoid the predicted failure by interventions of appropriate nature. If one can determine with interest tests that a particular youth is unlikely to find satisfaction in a program designed to teach culinary skills, it should be possible to help the student find a more suitable area for training. If one can predict from a reading test that a student will be unable to read a high school chemistry book, then remedial or alternative approaches can be tried. No sentient educator or psychologist uses tests for the test itself, because tests are for the purpose of answering questions that the educator or psychologist has a need to answer. Such questions may be related to the effectiveness of a program, for prediction of performance, for determining areas where remediation is needed, for monitoring individual progress. Tests can be highly useful tools if they have good validity; they may be useless or perhaps even harmful when wrongly used.

Because some tests have shown themselves to be useful in educational settings, they were well accepted. Eventually, considerable dependence was placed on the performance of students on tests; we came to expect them to do more than they were ever designed to do and many new tests were developed without adequate care. Thus, during the '50s and '60s and until today, an anti-testing movement developed. Several books that castigated test users and developers appeared on the market and some tests used for personnel selection were even denounced on the floor of the U. S. Senate. Much of the criticism

was earned by inexperienced or careless individuals who were not sophisticated in test use. Buying a test, reading the manual and administering the test seemed to some persons to be appropriate behavior. Such individuals, often with very good intentions, were at least in part, responsible for the misuses of tests that well trained professionals deplored. Naive individuals tended to expect more of a test than it was ever designed to offer. For example, part of the heavy criticism of intelligence tests included complaints that they did not provide information on motivation for learning; while an experienced clinical psychologist may make inferences about motivation from performance on tests, it was more likely that the psychologist would use a variety of sources in determining probable motivation and that the psychologist would be well aware of the variability of motivations under differing circumstances.

Tests are tools, just as weight scales and stethoscopes and volt meters are tools. They provide information that can be helpful in planning and evaluating, but they do not make decisions. Decisions are human prerogatives. (Recent developments of computers have indicated the usefulness of those tools in the decision-making process, but computers only do what they are programmed to do by those who decide to program them, and so the ultimate responsibility is with humans still.)

Selection of Measurement Tools

When one has a question that one has reason to believe one can answer more accurately or more rapidly by using a test, one can select a measure appropriate for the purpose. How the particular measure is selected will depend on several things. One obvious limitation is the skill one has in using a particular measure. For example, the question may be one of determining neurological impairment and one may wish to use one of the batteries developed for such evaluation, for example, the Reitan battery. If one is not trained in the use of that

battery, obviously it cannot be the choice. If a teacher has reason to believe that one of the children in the classroom is mentally retarded, then an individually administered test of intelligence would be helpful in clarifying the question; however, that teacher may have been given training in using criterion referenced tests and achievements tests; choice for using the Wechsler Intelligence Scale for Children-Revised is not an option. The teacher has the option of seeking the services of a psychologist trained in using that test, but not of borrowing a WISC-R kit and administering the test, or of deciding to determine classification of mental retardation on the basis of some test developed for screening purposes. Selection of tests that one may use personally depends in part on training and experience with the particular test and with test practices in general.

Some tests do not require extensive training. However, those who use standardized measures should have familiarity with the general processes of test construction and of the general procedures of test administration, scoring and interpretation. Specific procedures for a test are ordinarily given in the test manual. Information about the uses and usefulness of various tests can be gained by reference to one of the many textbooks in the field and from the series of Mental Measurements Yearbocks edited during the past few decades by Oscar H. Buros; these volumes are usually available in university libraries and in many evaluation centers. Reference to them can aid one in selection of an appropriate measure.

When two or more tests seem to be about equally useful for determining information needed, it is not uncommon for those who use tests to choose simply on the basis of personal preference. Nationwide surveys have indicated that some tests are much more "popular" than others. Why this is so is something of a mystery. Perhaps the more often used tests have wider scope, have demonstrated

better validity, or have greater efficiency. For many of the very commonly used tests, one or more of these reasons is probably correct. However, there are a few rather popular tests that do not have good reliability or validity and so their appeal must lie elsewhere. It is to be hoped that the appeal lies not in attractive packaging or good salesmanship.

III. SELECTED VOCATIONAL MEASURES

Introduction

The measures listed in this section were selected from a list of 108 measures originally chosen for review because they appeared to have potential for use with persons enrolled in regular vocational programs in regional schools. The criteria for initial review were (1) potential for use in curriculum development evaluation for the handicapped, (2) potential use in the individual assessment, and (3) appropriate grade-age level. The criteria for selection for annotation were (1) reviews suggested potential usefulness for curriculum evaluation or assessment (prediction for vocational training), (2) acceptable reported reliabilities/validity, and (3) features of test indicating appropriateness for use with handicapped persons (e.g., designed for non-readers; for deaf students, etc.).

Abbreviations are minimal in the sections below and many are self-explanatory. However, the following should be noted: (1) "K-R" refers to the Kuder-Richardson formulas for estimating reliability (see Guilford, 1965), and (2) ".82," ".70," refers to a correlation coefficient, usually Pearson r (Roscoe, 1969). The tests, for convenience, have been categorized by type.

A. SPECIFIC OCCUPATIONS

(1) Computer Programming

Title: Computer Programmer Aptitude Battery

Author: Palormo, J. M.

Publisher: Science Research Associates, Inc.

Level: Young adults (applicants for training or employment as a computer programmer).

Description: Six scores: (1) Verbal Meaning, (2) Reasoning, (3) Letter Series, (4) Number Ability, (5) Diagramming, and (6) Total Score. For use with applicants for computer training or employment. Correlations of total test score with grades in computer training range from .46 to .71. Tasks and tests have close correspondence to the tasks of programming.

Reliability: .67 - .94 (Split half, K-R 20) .95 (est. total).

(2) Engineering

Title: National Engineering Aptitude Search Test, The Junior Engineering Technical Society

Author: See Differential Aptitude Test.

Publisher: The Psychological Corporation

Level: Grades 9-12

Description: Tests are administered each spring at chapter centers of the Junior Engineering Technical Society. Five scores: (1) Verbal, (2) Numerical, (3) Science, (4) Mechanical Comprehension, (5) Total. (This is a special reprint of corresponding parts of the Differential Aptitude Test, Form M).

Reliability: See Differential Aptitude Test.

(3) Sales

Title: Detroit Retail Selling Inventory

Authors: Baker, H. J. & Bolker, P. H.

Publisher: Public School Publishing Co.

Level: Secondary students and young adults.

Description: Five scores: (1) Personality, (2) Intelligence, (3) Checking, (4) Arithmetic, and (5) Total Score. For use with students in secondary schools who are considering retail sales training and for inexperienced applicants for retail selling jobs in business.

Reliability: .76 - .98

(3) Sales, cont'd.

Title: Test for Ability to Sell: George Washington University Series

Author: Moss, F. A., Wile, H., Loman, William, Middleton, W., Hunt, T., George R., Schnell, W.

Publisher: Center for Psychological Service, George Washington University

Level: Ages 7 - 16, Adults.

Description: Tests ability to sell regardless of product. Subtests: (1) Social Consciousness, (2) Arithmetical Ability, (3) Memory, and (4) Ability to Follow Directions. (A validity coefficient on ability to select a good salesman of .54(N=100) is reported.)

Reliability: .91 (test - retest)

B. APTITUDE BATTERIES

Title: Differential Aptitude Test, 2nd edition, 1962

Author: Bennett, G. K., Seashore, H. G., and Wesman, A.G.

Publisher: Psychological Corporation

Level: Grades 8-12 and Adults.

Description: Consists of 8 tests: (1) Verbal Reasoning, (2) Numerical Ability, (3) Abstract Reasoning, (4) Clerical Speed and Accuracy, (5) Mechanical Reasoning, (6) Space Relations, (7) Language Uses: Spelling, and (8) Language Uses: Grammar. Norms available on approximately 50,000 boys and girls representing a cross section of the U.S.

Reliability: .87 - .94 (except for Mechanical Reasoning which, for girls, is reported as low as .70.

Title: Flanagan Aptitude Classification Tests

Author: John C. Flanagan

Publisher: Science Research Associates, Inc.

Level: First edition, Grades 10-12, Adults; second edition, Grades 9-12.

Description: Two editions. I. First edition has 16 tests: (1) Inspection, (2) Coding, (3) Memory, (4) Precision, (5) Assembly, (6) Scales, (7) Coordination, (8) Judgment/Comprehension, (9) Arithmetic, (10) Patterns, (11) Components, (2) Tables, (13) Mechanics, (14) Expression, (15) Reasoning, (16) Ingenuity.
II. Nineteen test edition is the same as the 16 test edition, except that it adds (17) Vocabulary, (18) Planning, and (19) Alertness. Battery of tests intended for use in counseling high school students in vocational education planning. Designed for predictions of success in 37 occupational areas of which 17 do not require college preparation. Time for administration of the battery is approximately 10½ hours.

Reliability: Range for all tests, .50 - .90.

Title: General Aptitude Test Battery

Author: None

Publisher: U.S. Employment Service, Government Printing Office

Level: Grade 12 and adults.

Description: Nine factors: (1) Intelligence, (2) Verbal Aptitude, (3) Numerical Aptitude, (4) Spatial Aptitude, (5) Form Perception (6) Clerical Perception, (7) Motor Coordination, (8) Finger Dexterity, (9) Manual Dexterity. Raw scores can be converted to percentile ranks or standard scores; standard scores can be compared with over 36 occupational aptitude patterns in over 800 jobs; administration time = 2½ hours.

Reliability: .80 - .90 (test-retest, parallel forms).

Title: Guilford-Zimmerman Aptitude Survey

Authors: Guilford, J. P. and Zimmerman, W. S.

Publisher: Sheridan Psychological Services, Inc.

Level: Ages 9-16 and Adults.

Description: Seven parts: (1) Verbal Comprehension, (2) General Reasoning, (3) Numerical Operations, (4) Perceptual Speed, (5) Spatial Orientation, (6) Spatial Visualization, and (7) Mechanical Knowledge. If used for vocational guidance, authors recommend the use of the entire battery which requires approximately 3 hours.

Reliability: .89 - .96 (internal consistency)
.88, .92, and .92 (split-half)

Title:	Non-Reading Aptitude Test Battery
Author:	None
Publisher:	U.S. Employment Service Government Printing Office
Level:	Grades 9-12, Adults (disadvantaged)
Description:	This test is a non-reading adaptation of the general aptitude test battery. Has 14 tests, 9 scores: (1) Numerical, (2) Intelligence, (3) Verbal, (4) Spatial, (5) Clerical, (6) Perception, (7) Form Perception, (8) Finger Dexterity, (9) Motor Coordination (10 paper and pencil tests and 4 performance tests).
Reliability:	For research purposes only.

C. MECHANICAL ABILITY

Title:	Bennett Mechanical Comprehension Test
Author:	Bennett, G. K. and Owens, W. A.
Publisher:	Psychological Corporation
Level:	Grades 9-12, Adults.
Description:	A revision of the Test of Mechanical Comprehension, Forms AA, BB, W1. Designed to measure the ability to understand and perceive the relationship of physical forces and mechanical elements in practical situations.
Reliability:	.81 - .93 (internal consistency).

Title:	Chriswell Structural Dexterity Test
Author:	Chriswell, M. I.
Publisher:	C. H. Stoelting Co., Vocational Guidance Service
Level:	Grades 8 and 9.
Description:	Designed to measure structural dexterity which is defined as ability to translate the visualization of structures into specific motor responses. This is analogous, or closely related to interpreting blueprints through appropriate manual bench and machine work.
Reliability:	.86 (test-retest)

Title: Mechanical Reasoning: Differential Aptitude Test

Authors: Bennett, G. K., Seashore, H. G., and Wesman, A. G.

Publisher: Psychological Corporation

Level: Grades 8-12, Adults.

Description: See Differential Aptitude Test.

Reliability: See Differential Aptitude Test.

Title: Space Relations: Differential Aptitude Test

Author: Bennett, G. K., Seashore, H. G., and Wesman, A. G.

Publisher: Psychological Corporation

Level: Ages 10:0 - 13:11; 15:0 - 18:0

Description: See Differential Aptitude Test.

Reliability: See Differential Aptitude Test.

D. MANUAL DEXTERITY

Title: Crawford Small Parts Dexterity Tests

Author: Crawford, J. E. and Crawford, J. M.

Publisher: Psychological Corporation

Level: High School and Adults

Description: Two scores: (1) pins and collars, and (2) screws. Designed to measure fine eye-hand coordination as related to assembly and adjustment of such devices as electronic equipment, electric clocks, hearing aids, and the manipulation of small hand tools. Low correlation between parts 1 and 2 (.10 to .50). Norms are provided for several groups: students, job applicants, and employees.

Reliability: Part 1: .80 - .91
Part 2: .90 - .95

Title: Hand Tool Dexterity Test

Author: Bennett, G. K.

Publisher: Psychological Corporation

Level: Adolescents and adults.

Description: A work sample test designed to measure proficiency in the use of ordinary tools such as a screwdriver or wrench. Norms available on a large sample of males from a paper factory.

Reliability: .91 (test-retest).

Title:	Purdue Pegboard
Author	Tiffin, J. (Purdue Research Foundation)
Publisher:	Science Research Associates
Level	Ages 9-16, adults.
Description:	Five scores: (1) Left Hand, (2) Right Hand, (3) Both Hands, (4) Right plus Left, plus Both Hands, (5) Assembly. Designed to aid in selecting employees for industrial jobs requiring manipulative dexterity. Purports to measure gross movements of both fingers and hands and arms and tip of the finger dexterity. (5-10 minutes administration time.)
Reliability:	1. .60 - .76 (test-retest) 2. Spearman Brown formula reliabilities for 3 trial scores range from .82 to .91.

E. INTERESTS

Title:	Geist Picture Interest Inventory
Author:	Geist, Harold
Publisher:	Western Psychological Services
Level	Ages 8-16, Adult.
Description:	Eighteen scales for males and 19 scales for females; 12 interest scores: (1) Clerical, (2) Mechanical, (3) Musical, (4) Scientific, (5) Outdoor, (6) Persuasive, (7) Literary, (8) Computational, (9) Artistic, (10) Social Service, (11) Dramatic, (12) Personal Service (female only). Seven motivation scores: (1) Family, (2) Prestige, (3) Intrinsic and Personality, (4) Environmental, (5) Financial, (6) Past Experience, (7) Could Not Say. Test is a pictorial approach for young people and others who have difficulty with verbal materials.
Reliability:	.57 - .73 (mdn. r, test-retest, N=1659)

Title:	Geist Picture Inventory: Deaf Form: Male
Author:	Geist, Harold
Publisher:	Western Psychological Services
Level:	Ages 7-16 and Adults.
Description:	For deaf and hard of hearing males. Adaptation of the Geist Picture Inventory, 10 scores: (1) Clerical, (2) Mechanical, (3) Persuasive, (4) Outdoor, (5) Scientific, (6) Literary, (7) Artistic, (8) Computational, (9) Social Service, (10) Dramatic.
Reliability:	.53 - .57

Title: Gordon Occupational Checklist

Author: Gordon, L. V.

Publisher: Harcourt-Brace Jovanovich, Inc.

Level: High School Students.

Description: Designed for use with students not planning to enter college. Five or 11 scores: (1) Arts, (2) Outdoors, (3) Technology, (4) Service, (5) Business; plus 6 optional summary scores using the preceding 5 scores and a total.

Reliability: High to middle .80's.

Title: Hall Occupational Orientation Inventory, 2nd Edition

Author: Hall, L. G., Tarrier, R. V., Shappell, D. L.

Publisher: Scholastic Testing Service, Inc.

Level: Age 16 and Adults.

Description: Scales are directional. Designed to measure a person's orientation away from or towards areas of interest described by the scales. The 2nd edition inventory yields 22 scores: (1) Risk, (2) Information-Knowledge, (3) Independence-Creativity, (4) Belongingness, (5) Security, (6) Aspiration, (7) Self-Actualization, (8) Esteem, (9) Routine-Dependence, (10) Things Orientation, (11) Data Orientation, (12) People Orientation, (13) Aptitude Concern, (14) Location Concern, (15) Physical Abilities Concern, (16) Monetary Concern, (17) Co-Worker Concern, (18) Environment Concern, (19) Qualifications Concern, (20) Time Concern, (21) Defensiveness, (22) Personal Satisfaction. Recommend use as a research instrument only.

Reliability: .84 (Mdn. r for original 23 scales, N=1400).

Title: Handerson Analysis of Interest, 2nd Edition

Author: Henderson, Robert W.

Publisher: Robert W. Henderson, 948 St. N.E., Massillon, Ohio

Level: Ages 9-16, Adults.

Description: Measures occupational preferences in 14 areas: (1) Clerical, (2) Business Service, (3) Statistics and Accounting, (4) Persuasive, (5) Managerial, (6) Physical Science, (7) Biological Science, (8) Engineering, (9) Social Science, (10) Art and Music, (11) Teaching, (12) Writing, (13) Manual, and (14) Mechanical. (Correlates from .30 to .77 with Kuder.)

Reliability: .80 - .96.

Title: Inventory of Vocational Interest: Acorn National Aptitude Test

Author: Kobal, A., Wrightstone, J. W., and Kunze, K. R.

Publisher: Psycho-Metric Affiliates

Level: Ages 7-16, Adults.

Description: Five scores: (1) Academic, (2) Artistic, (3) Mechanical, (4) Business-Economic, (5) Farm-Agricultural. Short time-span for testing, simple scoring, written in the language of the students, well-balanced occupations for social and prestige appeal.

Reliability: "Inadequate," recommend for research purposes only.

Title: Kuder General Interest Survey

Author: Kuder, G. F.

Published: Science Research Associates, Inc.

Level: Grades 6-12.

Description: This is an extension of the Kuder Preference Record-Vocational Form C. Eleven scores: (1) Mechanical, (2) Outdoor, (3) Scientific, (4) Computational, (5) Artistic, (6) Persuasive, (7) Literary, (8) Musical, (9) Clerical, (10) Social Service, and (11) Verification. This version of the Kuder Preference Record-Vocational, Form C was developed for use with junior high school level children and is written at a sixth grade reading level.

Reliability: Above .70 (K-R-20).

Title: Kuder Preference Record-Vocational (Form C)

Author: Kuder, G. F.

Publisher: Science Research Associates, Inc.

Level: Grades 9-12 and Adults.

Description: Ten interest scales: (1) Mechanical, (2) Outdoor, (3) Scientific, (4) Computational, (5) Artistic, (6) Persuasive, (7) Literary, (8) Musical, (9) Clerical, (10) Social Service. Also provides a score on (11) Verification. Offers percentile ranks for males and females. Stanines for women and men in 41 occupational clusters.

Reliability: .47 - .75 (test-retest)
K-R for 10 scales in .70's and .80's.

Title: Minnesota Vocational Interest Inventory

Author: Clark, K. E. and Campbell, D. P.

Publisher: Psychological Corporation

Level: Males, age 15 and over

Description: Designed to measure interest in males who are not planning to attend college. (I) 21 occupational scales: (1) Baker, (2) Food Service Manager, (3) Milk Wagon Driver, (4) Retail, (5) Stock Clerk, (6) Printer, (7) Tabulating Machine Operator, (8) Warehouseman, (9) Hospital Attendant, (10) Pressman, (11) Carpenter, (12) Painter, (13) Plasterer, (14) Truck Driver, (15) Truck Mechanic, (16) Industrial Vocational Teacher, (17) Sheet Metal Worker, (18) Plumber, (19) Machinist, (20) Electrician, (21) Radio-TV Repairman. (II) Nine area scales: (22) Health Service, (23) Mechanical, (24) Office Work, (25) Electronics, (26) Food Service, (27) Carpenter, (28) Sales Office, (29) Clean Hands, and (30) Outdoors. Inventory was designed to measure the similarities between the student taking the interest survey and those of men employed in a variety of non-professional occupations.

Reliability: .62 - .82 (test-retest).

Title: Ohio Vocational Interest Survey

Author: D'Costa, A. G., Winefordner, D. W., Odgers, J. G. and Koons, P. B., Jr.

Publisher: Harcourt-Brace Jovanovich, Inc.

Level: Grades 8-12.

Description: Cannot be scored locally. Twenty-four scores: (1) Machine Work, (2) Manual Work, (3) Personal Services, (4) Caring for People or Animals, (5) Inspecting and Testing, (6) Clerical Work, (7) Crafts and Precise Operations, (8) Customer Services, (9) Skilled Personnel Services, (10) Nursing and Related Technical Services, (11) Training, (12) Numerical, (13) Literary, (14) Appraisal, (15) Applied Technology, (16) Agriculture, (17) Promotion and Communication, (18) Supervision and Management, (19) Artistic, (20) Music, (21) Entertainment and Performing Arts, (22) Sales Representative, (23) Teaching-Counseling-Social Work, (24) Medical. National standardized sample of 45,000 students, grades 8-12.

Reliability: .80 (test-retest).

Title: Pictorial Interest Inventory

Author: Scarborough, B. B.

Publisher: Scarborough, B. B.

Level: Adult males, non-readers or poor readers.

Description: For research purposes only. Eleven scores: (1) Personal Service, (2) Protective and Custodial, (3) Clerical and Sales, (4) Mechanical, (5) Farming, (6) Building and Maintenance, (7) Vehicle Operators, (8) Skilled-Sedentary, (9) Electrical Workers, (10) Assembly Line Workers, and (11) Natural Processors.

Reliability: For research purposes only.

Title:	Picture Interest Inventory
Author:	Weingartan, K. B.
Publisher:	CTB/McGraw-Hill
Level:	Grade 7 and over.
Description:	Nine scores: (1) Interpersonal Service, (2) Mechanical, (3) Natural, (4) Business, (5) Esthetic, (6) Scientific, (7) Verbal, (8) Computational, (9) Time Perspective. Non-verbal inventory. Closely related to the Occupational Interest Inventory.
Reliability:	.69 - .92 (test-retest).

Title:	Reading Free Interest Inventory
Author:	Becker, R. L.
Publisher:	American Association on Mental Deficiency
Level:	Adolescents and Adult educably mentally retarded.
Description:	Designed as a non-reading measure of vocational preference for educably mentally retarded youth in specific job areas. Job areas selected on basis of realistic potential for EMR. Eleven categories for males and eight for females.
Reliability:	Males: .68 - .92 (K-R-20); .75 - .92 (test-retest). Females: .69 - .96 (K-R-20); .72 - .88 (test-retest).

Title:	Strong Vocational Interest Blank for Men
Author:	Strong, Edward K., Jr., Campbell, D. P. Birdie, R. F. and Clarke, K. E.
Publisher:	Stanford University Press
Level:	Ages 16 and over.
Description:	Eighty-four scales: (1) 22 basic interest scales, (2) 54 occupational scales, (3) 8 non-occupational. Also, 6 administrative indices.
Reliability:	.90 (test-retest) (30 days). .80 (mdn. odd-even).

Title:	Strong Vocational Interest Blank for Women
Author:	Strong, E. K., Jr., and Campbell, D. P.
Publisher:	Stanford University Press
Level:	Ages 16 and over.
Description:	Eighty-one scales: (1) 19 on basic interest, (2) 58 on occupational and (3) 4 non-occupational. Also, 6 administrative indices.
Reliability:	.68 (test-retest).

F. MISCELLANEOUS

Title: Career Maturity Inventory

Author: Crites, J. O.

Publisher: CTB-McGraw-Hill

Level: Grades 6-12.

Description: Consists of (I) Attitude scale and (II) five competence tests in a research edition. The five competence tests are (1) Self-Appraisal, (2) Occupational Information, (3) Goal Selection, (4) Planning, (5) Problem-Solving.

Reliability: (I) Attitude Scale - .74 (K-R-20).
(II) Competence Test - .72 - .90 (with 2 exceptions). Two low coefficients were reported for "problem-solving" subtest in the sixth and the seventh grades.

Title: Dailey Vocational Test

Authors: Dailey, J. D. and Hoyt, K. B.

Publisher: Houghton-Mifflin

Level: Grades 8-12 and Adults.

Description: Consists of (1) Technical and Scholastics Test, (2) Special Visualization Test, and (3) Business English Test. Extensive norms available.

Reliability: TST: .84 - .96
SVT: .84 - .93
BET: .90 - .95

Title: General Clerical Test

Authors: Hay, E. N., Hunt, T., Katzell, R. A. and Wonderlic, E. F.

Publisher: Psychological Corporation

Level: Ages 9-16.

Description: Formerly called Psychological Corporation General Clerical Test. Four scores: (1) Clerical Speed and Accuracy, (2) Numerical Ability, (3) Verbal Facility, and (4) Total Score.

Reliability: .94.

Title: McGuire Safe Driver Test Scale and Interview Guide

Author: McGuire, Frederick R.

Publisher: Western Psychological Services

Level: Driving age.

Description: For use with prospective motor vehicle operators. Seventy-eight temperament, values and interest items. Purpose of the test is to help select safe drivers. Six validity studies found that the test can discriminate between better and poorer drivers, but only when both are tested after already being hired.

Reliability: .88.

Title:	Wilson Driver Selection Test
Authors:	Wilson, C. L. and Bruce, M. M.
Publishers:	C. L. Wilson and M. M. Bruce, Ph.D., Publishers
Level:	Driving age.
Description:	Consists of six chiefly non-verbal aptitude tests and an unscored page of bibliographical information. Purpose of the test is to determine ability to operate motor vehicles with minimum risk. Six sub-tests are: (1) Depth Visualization, (2) Visual Attention, (3) Recognition of Simple Detail, (4) Recognition of Complex Detail, (5) Steadiness, (6) Eye-Hand Coordination.
Reliability:	.59 - .93; total battery: .84.

IV. TESTS CONSIDERED FOR PROJECT VITA

One of the purposes of this project was to see whether a battery of tests suitable for use with handicapped vocational school students could be developed so that guidance of students would be facilitated. The procedure adopted for this task was first to survey the literature on use of individual tests (and some recommended but not validated batteries suggested). Much of the experimental work found was done with non-handicapped populations. That done with handicapped persons was for the most part done with mentally retarded individuals and the level of retardation of populations studied was frequently not given. Results of different studies were not in agreement. Some studies suggested that "normal" students in their teens were not, as a group, very firm in their occupational choice or consistent in their performance in occupational training classes. Furthermore, tests that were reported to be relatively good predictors (for normal secondary level students) did not necessarily serve as good predictors of later job performance, at least as that performance was measured by the particular criteria selected as indicating success. Several reports in professional journals provided suggested batteries that the writers deemed appropriate for a particular handicapping condition, but even these did not offer rationale for selection of the battery, even though the rationale could be inferred from the names of the tests recommended. The failure to find a battery that had been demonstrated to be successful with vocational school students who might have varying handicaps led the project staff to study the potential of numerous tests to determine which ones might be suitable for several types of handicapped students.

In a number of studies reported, the level of intelligence was indicated as a factor in successful job performance as well as in training programs.

Some studies did not show predictive validity for intelligence tests; a number of these had used a relatively narrow range of intelligence, such as IQ between 50 and 75, and so it was perhaps due to that factor that results were negative. One of the most large scale studies (10,000 Careers) did find relationships between occupation and IQ determined prior to entry into the occupation, but there was IQ overlap in some of the different occupational groups. Studies of different professional and skilled groups have indicated that probably there are "minimal" levels of intelligence needed to enter certain fields. In addition, of all the tests available to would-be test users, the individually administered tests of intelligence have probably been the most consistently related to other variables. Therefore, it was decided that an intelligence test should be included in the battery. Review of tests available and of investigations of use of tests clearly indicated that the most likely useful test would be the Wechsler Scale appropriate for the age of the student. Because of restrictions on testing in many schools and the difficulty of getting consent for using an intelligence test with students in an investigation not directly related to that student's own welfare, it was not possible to have project staff administer these tests. In the process of evaluation for certification as a student with "special needs," many of the students had been tested with a Wechsler Scale recently and so it was decided that these results were to be used in the project. Data for students who had been given other intelligence tests were not used, and those students were simply excluded from the statistical analyses where intelligence test results were needed.

Since the development of the Strong Vocational Test in the late 1920's, interests have been believed to be important for planning vocational training. Large numbers of investigations have indicated the usefulness of such tests. The earlier interest tests (such as the Strong) were developed by having persons

who were successful in various occupations take a test composed of many items on which they could express preference. Statistical analyses then determined on which items there was difference in groups from different professions. Thus, the tests did not predict ability, but interests similar to those already in a particular line of work. Later interest tests have been developed in other ways, sometimes by "armchair" analysis or by seeking face validity or construct validity.

The interest tests used with non-handicapped populations had two general limitations for this project. The major one is that the reading level of such tests as the Strong Vocational Interest is simply much too high for a very large proportion of the handicapped students in this population; indeed, a substantial proportion of these students had been identified as "learning disabled," meaning in part that their reading skills were far below that expected for age and grade placement levels. Of the tests with no reading, one which seemed to have potential was the Reading Free Interest Inventory, a measure developed for use with mentally retarded students; for our purposes, a problem with that measure, however, was that it is primarily concerned with unskilled and semi-skilled occupations and thus did not have sufficient occupational range for measuring a population of students who ranged in occupational potential up through at least the skilled and technical levels.

After review of approximately 20 interest measures, the decision was made to use two, the Geist Picture Interest Inventory and another measure that contained both interest and personality sub-scales and which will be discussed later. The Geist has 11 sub-scales and offers relatively wide scores; it is somewhat similar to the Kuder Interest Measures in that the scales are reported in terms of general areas of interest (e.g., scientific, mechanical, social service) rather than in terms of specific occupations as the Strong does. This

appeared to be another advantage for the purposes of this project. In view of the choice of these students to attend a vocational school where they could engage in "hands on" activities and practical training for skilled work, one might hypothesize that many of them would obtain relatively high scores on such sub-scales as mechanical and computational, but perhaps lower scores indicating lesser interest in literary and musical areas. The test has acceptable reliability and validity coefficients.

Motor skills seemed to be an area which should be included in the battery because of the motoric nature of much of the work for which these students were training. Many of the training programs in vocational schools involve the use of the hands, often with fine motor coordination required. Several motor tests were considered for inclusion in the battery. For example, the Oseretsky Test of Motor Skills has a wide range of motor skills measured and the ceiling (top of range) is clearly sufficiently high for the purpose. However, it requires a time-consuming individual administration and the norms available are on populations that are not typical of high school students in this country. Other tests considered used primarily paper-and-pencil and were eliminated for that reason, or were measures of gross motor skills. Of the measures that seemed to be likely to "tap" hand and finger dexterity and to be of reasonable length, the most promising appeared to be the Purdue Pegboard, a test on which a large body of supportive research was available.

The fourth area for which a measure seemed advisable was personality. Since a major aim of the project was to develop a battery that could be administered to large numbers of students in reasonably brief time if it had potential for widespread use, none of the individually administered projective measures appeared appropriate. Another difficulty in using such devices is that they require a high level of training and experience. Although there were on the

project staff individuals who could have done this work, to include such tests would limit the usefulness of a battery. However, many of the reports in the professional journals clearly indicate that personality factors are of great importance in maintaining a job; those who are discharged for cause are not usually those who cannot perform the task; they are those whose social skills are inadequate or inappropriate. The paper-and-pencil measures of personality were deemed inappropriate for this population for several reasons, one of which is that the test items were too difficult for most of these students to read. Various measures of individual traits, many developed for experimental purposes, were considered and not selected because of narrow range. Other measures, such as the Minnesota Multiphasic Personality Inventory, were neither readable by these students nor suitable in content provided. The California Personality Inventory was also of too high a reading level. Another problem encountered was the difficulty of obtaining consent from those to whom one would administer a personality inventory for research purposes. A compromise decision was to use the Holland Vocational Preference Inventory, a measure that is intended to measure both interests and certain personality variables that appeared relevant for occupational success.

A wide variety of other measures were considered. Final decisions depended on reported usefulness of the tests, potentially unique contribution to the battery, areas of information that have been demonstrated to be related to job success, and feasibility for getting cooperation in this project and in future users of any battery that was helpful in predicting.

In addition to these formal measures, rating scales were developed to gather information on certain work and personal habits that have been reported to be related to work success. These scales are described in detail in the next section of this report.

V. METHODOLOGY AND RESULTS

Introduction

The focus of this portion of Project VITA was to determine whether a battery of tests used with a sample of handicapped students enrolled in vocational education program(s) could aid in predicting success in vocational training. Procedures, instruments, and student samples were selected to provide data on: (1) interest measures, (2) aptitude measures, (3) short-term experiences within specific training programs; the sample consisted of groups of students from two different vocational-technical high schools. For this purpose, only students in the ninth and tenth grades were used because after those levels, students are scheduled into programs that are no longer exploratory. Primary focus of the work was with students at the ninth grade level; tenth grade level students were added in order to determine how these students might differ from ninth grade ones and the degree to which teacher ratings for completed work was related to test scores.

Characteristics of the sample are shown in Table V-1. For some measures, it was not possible to get complete data on all measures because of absences, refusals, or unavailability of teachers to make ratings. Numbers of subjects for which analyses were done for each variable are reported in the tables showing data.

TABLE V-1

Characteristics of Sample

N=75

Category	School I (N=37)	School II (N=38)	Totals
Males	34	30	64
Females	3	8	11
Grade 9	22	38	60
Grade 10	15	0	15
Age Mean Range	14-18	14-18	14-18
Educational Classification			
LD	30	21	41
MR*	2	12	14
ED	3	0	3
Other**	2	5	7

*includes MR who are physically handicapped.
**includes hearing impaired, speech impaired and physically handicapped persons who are not MR.

Measures Selected:

Criterion for Evaluation of Measures. The primary criterion against which measures were evaluated was the teacher's rating of the student's rank in the class in which the student was working at the time of the data gathering; for these students, data was gathered from teachers near the end of the eight-week units; thus, grades for the unit were not available. It was expected that student ranking in class would provide a broader range of scores than the A through F grading scheme and directions given teachers for this ranking were designed to de-emphasize some of the factors (such as making allowance for the handicap) that teachers may take into account in assigning grades. The instructions were:

> "Think of the major projects that this student worked on in your shop/lab. Now, please rate the student on the quality of work done in those projects. Compare the student with others in the shop/lab."

A line with centile rankings from the Bottom 10% to the Top 10% was provided; and teachers checked the appropriate centile area for each student. At each school, the Director of Special Education collected this information from teachers; the procedure was explained to teachers and questions from teachers about the project were answered only to the extent necessary to obtain ratings. Since teachers were familiar with rating scales, little explanation was needed. It would have been advantageous to have grades assigned to the units as a check against these ratings, but these could not be given until the end of the unit and so no analyses of grades on these units was possible.

Ability Measures. Intelligence: For the majority of these disabled students, the school's psychologists had given individual intelligence tests prior to the student's admission into the program. In this state, a psychological evaluation is required as a part of the "core evaluation" for determination of

eligibility for special education. The Wechsler Intelligence Scales were ordinarily used and data from these measures were available for most students. (For deaf students, only the Performance Scales were used in most cases because of the difficulty in administering the Verbal sub-tests to deaf persons.) Review of literature of predictions of vocational success have often used one of the Wechsler Scales and when it has been used in combination with other measures, it has often been shown to have greater predictive power than many other measures. Rather than using a group intelligence test, which would be a major problem with some of these students who have somewhat limited reading ability (which tests are less precise than individually administered tests), the data from Wechsler test results were included in analyses. It is recognized that the use of data that were not collected under conditions under the control of the project staff may have unknown sources of error; however, the scope of the project did not permit readministration of individual tests. (Group tests were not considered likely to add much, if any, additional information, and all psychological examiners were trained and certified by the State Department of Education.)

Motor Skills. Of the tests reviewed, the most promising for the purpose appeared to be the Purdue Peg Board. This test can be used with a variety of disabilities (perhaps even blindness), is relatively easy to administer and score, takes a reasonable time for administration, has good reported test-retest reliability, and showed greater promise for prediction than many of the other measures available. If a battery is to be used, the characteristics listed above are important.

Interests. Two measures were chosen in this area. It was important to keep reading level low so that mentally retarded and learning disabled students could be tested; the Geist Picture Inventory test does not require reading and covers a wider range of occupations than, for example, the Reading Free Interest

Inventory that was developed for use with mentally retarded individuals. The Geist inventory was designed for use with both males and females, in contrast to some other measures that seem to be primarily male-oriented. Research reports on the use of the Geist inventory indicated that it was a promising tool.

The second interest inventory chosen was Holland's Vocational Preference Inventory. This measure is based on the assumption that interests are related to personality variables. It was anticipated that this measure might provide some indication of personality traits as they might relate to success in job training, as well as information about interests.

Rating Scale. Ratings by teachers for each student were made for the students in the units. After reviewing studies on job success, the staff decision was to seek ratings on seven characteristics of student work or behavior. The items were selected to represent variables that have been reported to have relevance to maintaining a job in the community. On these ratings, teachers were asked to rate students by checking what they believed to be the most representative level, comparing students with others in the class, and using a straight line which had the characteristic listed above and the two extremes of the dimension at either end. Scores were obtained by using a ruler to determine where the check appeared on the 100 millimeter line; for scoring purposes, the line was subdivided into ten sections, and scores ranged from 1 to 10. (Scores were assigned to the nearest number when they fell between numbers.)

The areas on which ratings were made and the phrases placed at the ends of the lines were as follows:

1. Punctuality: Always on Time to Frequently Late.
2. Reaction to Supervision: Works Well with Supervision to Works Poorly with Supervision.
3: Interaction with Other Students in Shop/Lab: Many Interpersonal Difficulties to Gets Along Well.

4. Care of Equipment, Tools, and Materials: Very Irresponsible to Very Responsible.
5. Consistency of Work Completion: Works Steadily to Complete Tasks to Inconsistent Work.
6. Work Rules and Regulations: Consistently Ignores to Consistently Observes.
7. Attendance in Shop/Lab: Excellent to Poor.

In addition to the formal assessment measure used, data were gathered on a number of other variables such as sex, educational classification of disability, and age.

Method

The tests were administered by the two Special Education Directors at their respective vocational-technical schools. One school is located about twenty miles south of Boston and serves seven towns, many of them "bedroom communities" for Boston. This school is considered one of the most outstanding vocational-technical schools in the country and has been serving handicapped students for several years; the program for students with special educational needs is well established. The students coming to this school are primarily middle class, with a substantial proportion of "working class" residents living in the communities served. Most of the towns in the collaborative are in the 5,000 to 20,000 population range, but in some cases the towns tend to blend together; all could be considered to be essentially suburban or small town communities.

The second school is located about [illegible] miles north of Boston. Like the first school, it has a remarkably fine physical plant. The program for students with special needs is well established, but has not been in operation as long as that of the first school. The communities in the collaborative

are essentially small towns in the 3,000 to 10,000 population range and the socio-economic levels of the population are "working class" and middle class. These schools were chosen because of their scope, their reputations for excellent programs, and the willingness of the administrative staff to cooperate; another advantage was the demonstrated level of skill of the two special education directors, both of whom were well trained in research methodology and assessment procedures, thus helping to ensure that the tests would be carefully administered.

Tests were administered to small groups of students during regular school hours, except for the Purdue Pegboard, which was individually administered. All test data were gathered toward the end of an eight-week exploratory training unit.

Results of Analysis of Data

In one of the schools, more than one teacher was asked to rate the same student, who was in more than one instructional unit. This made possible a check on agreement across raters and may be assumed to give some indication of the consistency of student work in different units.

TABLE V-2

Inter-rater Reliabilities for Seven Ratings and Class Rankings

Item	Inter-rater Reliabilities	Correlation Between Ratings and Class Rank
1. Punctuality	.48	.41
2. Supervision	.64	.59
3. Interpersonal	.56	.69
4. Tools, equipment	.49	.49
5. Work Completion	.64	.60
6. Rules/Regs.	.70	.70
7. Attendance	.55	.55

It appears that agreement across raters in School II, where more than a single rating was available on the students, was not especially high. This could be due to several factors; difference in student behaviors with different teachers, difference in teacher interpretation of the terms, or combinations of these and/or other factors. Visual inspection of the data suggests the following patterns: (1) Two raters gave the same student high ratings and two gave the student low ratings, (2) three of the four raters agree but the fourth rater used much higher or much lower ratings for a student. There appeared to be relatively high agreement between teacher ratings and the class ranking (in centile) assigned; if ratings of student rank in class is used as the criterion for "successful training," then the teacher ratings are better predictors than any of the other measures used. However, this may be due to some "halo" effects because teachers did both sets of ratings. It seems probable that these class rankings assigned to students will reflect grades assigned to the students; several studies have shown that assigned grades are related to ratings and vice versa.

The seven ratings for each student were subjected to factor analysis to determine if there were seven different factors or less than seven. This analysis yielded two factors, not seven. Factor 1 was made up of the following ratings: Reaction to Supervision, Care of Equipment, Tools, and Materials, Consistency of Work Completion, and Work Rules and Regulations. Thus, it appears that high ratings on one of these is related to high ratings on the others. Factor 1 seems to be related to Work Behaviors and was therefore called the "Work Behaviors" factor. Factor 2 was made up of these items: Punctuality, Interaction with Other Students in the Shop/Lab, and Attendance; Factor 2 appears to be composed of behaviors not directly related to the work activity itself and was therefore called "Personal Behaviors." (Since poor attendance, tardiness and getting into difficulties with fellow students can all reduce

time in work or training, it could have been called Presence/Absence, but to do so would be to make the assumption that teachers were rating on the interaction dimension primarily in terms of behaviors that involved some substantial amount of time in interpersonal difficulties and there was insufficient evidence to indicate that teachers so rated.)

Factor 1 was the strongest factor, accounting for 62 per cent of the variance in ratings. Factor 2 accounted for only 11 per cent.

As a predictor of class rank ratings, Factor 1 was strongest, suggesting that the class rank ratings were more strongly influenced by Work Behaviors than by Personal Behaviors.

For School I there was only one rating for each student. However, several different teachers made the ratings because students were in various classes. Thus, ratings at both schools have a confounding effect. Neither tasks nor environments were the same for the two schools.

At School II, where more than one rating was available for students, assignments in class rank were made by shop instructors (e.g., automotive mechanics) and some by teachers who had the students in academic classes in which the curriculum was related to the shop work; it may be that the variability in ratings (e.g., 60, 75, 85) may reflect differing performance in the hands-on shop work and the academic related activities in the classrooms. At School I, ratings were for shop work; there is a consistent tendency for the ratings at School I to be higher than in School II for the items of work related activities (Factor I); for these analyses the average ratings of teachers in School II were used and this may have been reflected in this finding. Ratings in School I were significantly lower than in School II on the items of attendance, interpersonal relationships, and punctuality. It may be that the inclusion of teachers of academic classes in the ratings of School II is reflected in these

findings; many of these students have consistently had difficulty in academic work and inclusion of ratings of work in academic classes could have lowered the combined ratings if shop (the only area rated in School I) was the highest rating in both schools. It may be that the greater freedom in shop work permits more opportunities for students to interact and thus get into difficulty with other students.

Individual teachers tended to be internally consistent in use of the rating scale for the seven items ($\alpha = .76$). From the factor structure information, it appears that they did tend to make at least two major discriminations within the scale (Factors 1 and 2).

Prior School Grades and Class Rankings Assigned:

An effort was made to determine the extent to which the grades obtained in previous courses were related to the rank in class ratings gathered for these students. These data were available for 69 of the total group. The correlation between prior grade assigned on a shop course (e.g., A, B, C, D, F) and the class rank assignment by the teacher of the current course was .357; this moderate positive correlation is statistically significant at the .01 level, but it suggests that prior grades in shop classes may be a relatively poor predictor of class rank assignment in a different shop since the overlap in variance is just under 20 per cent. Even so, the prior grade received showed a higher correlation with class rank assignment than with any other variable.

Interpretation

It can be concluded that the teachers did make discriminations across students, but apparently the behaviors were to some extent interrelated (Factors 1 and 2).

From these data, one can say that there are some differences in the ways in which teachers rate the same students (shown by inter-rater correlations for

School II); it seems very likely that shop work of a given student received higher ratings than academic work, or shop teachers tend to give higher ratings than academic teachers; the first of these hypotheses seems more likely than the second on the basis of inspection of the raw data.

Intelligence Test Results:

Comparisons by Schools

Analysis of variance comparing School I and School II on the intelligence and sex variables indicated that the mean Wechsler Full Scale IQ for School I was significantly higher than in School II ($F=5.84$, $p= .05$). For School I, the Full Scale IQ mean was 96.15, s.d. 13.56; for School II, Full Scale IQ was 83.91, s.d. 16.3. Although both schools are serving students below the national norm of 100, School II appeared to have a student population that is more toward the lower end of the normal curve. The School I mean is near the national norm; School II mean is about one standard deviation below the national norm. The difference between schools on Wechsler performance IQ is significant with School I having a mean of 102.10, s.d. of 15.88, and School II having a PIQ mean of 87.54, s.d. 19.56 ($F=5.02$, $p= .05$). In Verbal IQ (VIQ), for School I, the mean VIQ was 91.09, s.d. was 11.75; for School II, the mean VIQ was 90.79, s.d. was 12.87; these differences are also significantly different ($F=5.51$, $p= .05$).

Full Scale Wechsler scores were available for 48 students, Verbal IQ for 44, and PIQ for 46. In some cases, complete data were not available because students could take only one part of the test because of disability. In other cases, the schools did not have the data because the local feeder schools had not sent it to them. This was especially true in School I, for which these data were available for only 13 of the 42 students. Therefore, one cannot draw firm conclusions about the differences between schools.

Comparison by Sex:

For the total group, significantly higher intelligence test scores were obtained for males than females. These means and standard deviations were:

Males: FSIQ mean was 89.21, s.d. 14.56; PIQ mean was 92.65, s.d. 18.39; VIQ was 85.08, s.d. 12.47.

Females: FSIQ was 78.67, s.d. 20.30; PIQ was 84.33, s.d. 24.06; VIQ was 75.63, s.d. 14.78.

On both the FSIQ and the VIQ differences, analysis of variance indicated that these differences approached, but did not reach statistical difference at the .05 level of confidence. (For FSIQ differences, F was 3.17; for VIQ differences, F was 3.53; for PIQ differences, F was 1.31. Although the differences were not statistically significant, they were sufficiently great to suggest that for comparisons it may be wise to analyze data for each sex rather than to combine them.

The differences across schools in intelligence test results would suggest that data for the two schools should be analyzed separately when the size of the groups on whom data are available is significantly large.

Analyses of Predictions Using IQ as Co-variate:

Three-way analysis of co-variance was conducted with Full Scale IQ used as the co-variate. FSIQ was a predictor ($F=3.223$, $p<.07$). Sex differences were also significant; females were given higher class rankings when scores were adjusted for FSIQ ($F=13.17$, $p=.001$). These data suggest that female students were assigned higher class rankings than males, even though females were lower in FSIQ.

Combining the ratings for Factor 1 (Reaction to supervision, Care of equipment, tools and materials, Consistency of work completion, and Observance of rules and regulations), results indicate that FSIQ is a strong predictor of Factor 1 ratings ($F=6.19$, $p=.02$). Thus, the higher the IQ, the higher class

rank assigned. When adjustments were made for IQ, females still were higher in class rankings than males ($F=9.76$, $p= .02$). However, when IQ was held constant statistically, there were no significant differences between grades or schools in ratings on Factor 1.

For Factor 2 (Punctuality, Interaction with other students, and Attendance), again FSIQ was strongly predictive ($F=17.34$, $p= .001$); the only other significant difference was that teachers in School A rated their students as having better attendance. With FSIQ, controlled statistically, the same sex differences were found for Factor 2 as for Factor 1; females were rated significantly more often in the "good" direction on the items of Factor 2 than were males.

Type of Handicap and Ratings

Class Rank ratings for the students were compared on the various disability areas. Because of the small numbers in some of the students for whom more than one handicap was listed (e.g., Physical Handicap plus Mental Retardation), some groups were combined (e.g., PH + MR were classified under MR because it was listed first). Mean ranking on quality of work was highest for the students who had Speech, Hearing, or Physical Handicaps; second highest rankings were assigned to the Learning Disabled; third were the students with Mental Retardation (most educable); lowest class rank assignments were those diagnosed as emotionally disturbed. Students had been identified by the interdisciplinary team that recommended that the students be qualified for educational services as having the disabilities listed. Mean rankings were as follows:

1. Speech/Hearing/Physically Handicapped - 76.33
2. Learning Disabled - 69.60
3. Mentally Retarded - 56.0
4. Emotionally Disturbed - 36.71

The grand mean for all students was 64.34. There appeared to be a

great difference in reported perception of the work (as indicated by class rank assigned) between groups; the teachers appeared to think that those with sensory and physical handicapping conditions were doing quite well in class, as indicated by assignment to about the seventh decile, and that the disturbed students did very poorly, since these students were assigned class ranks far below average.

On the basis of these data, it appears that simply knowing the disability group to which a student is assigned can give a fairly gross prediction of the way in which teachers rank students as compared with other students in the shop course. Type of handicap correlations with class rank of students provided correlations of between .50 and .70. At a more refined level of grading or of handicapping condition, somewhat different results might obtain; on a clinical analysis basis, one would expect that the predictions would probably improve rather than be worse.

Interest and Motor Tests

In an effort to determine the degree to which the Purdue Pegboard, the Geist Picture Interest Inventory and the Holland Vocational Preference Inventory predicted prior grades and the class rankings assigned by teachers, a step-wise multiple regression analysis was done.

VPI:

Results indicated that the collection of sub-scales of the Vocational Preference Inventory (VPI) did not predict class rankings as well as either Full Scale Wechsler IQs, handicapping condition, or prior school grade on similar class units. The sub-scales of the VPI which had the best, although relatively poor predictive value were Enterprising, Self-control, and Infrequency. The Enterprising scale is one in which high scores are interpreted to indicate that an individual tends to be "dominant, sociable, cheerful, and adventurous." The Self-control scale interpretation suggests "habitual inhibition of impulses

to act out motivation, thinking or phantasy...careful, smooth;" high scorers are described as inhibited, constricted, passive, and responsible, whereas low scorers are described as impulsive and having a tendency to act out. The Infrequency scale is interpreted to mean that an individual has preferences for unpopular, feminine, low status occupations and high scores are said to be suggestive of "atypical vocational preferences" and perhaps self-deprecating attitudes about themselves. The strongest predictor of class rank was the Enterprising scale, with a multiple correlation of .25 and accounted for approximately 6 per cent of the variance. Adding scores on the Self-control sub-scale increased the multiple correlation to .34 and variance accounted for to approximately 12 per cent. Adding to these the results of the Infrequency scale increased the multiple correlation to .39 and the variance accounted for to approximately 15 per cent. Thus, while there is approximately a 15 per cent overlap, about 85 per cent of the variance is left to be accounted for by something other than the VPI. It is interesting that the three sub-scales that had some predictive value were among those which are less "interest" than personality measures. It is not particularly surprising that teachers might assign somewhat higher class ranks to students who were enterprising and showed self-control than to students who show less of these traits. The analysis of the VPI and combined seven ratings by teachers (e.g., punctuality + work rules, etc.) yielded results that were highly similar to those on class rankings, which is not especially surprising since the same teachers rated and ranked the students and the possibility of halo effects cannot be ruled out; the only interesting difference found was that adding the Infrequency scale tended to raise the multiple correlation obtained by adding Self-control to Enterprising from .24 to .37, suggesting that adding data from the Infrequency Scale was more useful for predicting teacher ratings than predicting teacher rankings of students in class.

When the items of Factor 1 of the ratings were used to predict class ranking, the results provided a multiple correlation of .75; adding Factor II increased the correlation to only .78. Thus, it appears that the teacher ratings on Factor 1 items were much more closely correlated with class ranking than VPI scale scores (i.e., .75 versus .35). Altogether, it appears that at least for situations such as the one for this study, the VPI is not a particularly useful measure. One must recognize, however, that with the small number of students involved in this study it was not possible to determine predictions for individual classes (units) and correlations might be improved if large samples of students in each of several class units were used. One must also keep in mind the fact that the criterion was teacher assignment of class ranking and the fact that the VPI is described by the developer as being best used as a "brief screening inventory." The results of these analyses tend to suggest that one may wish to be extremely cautious in drawing conclusions based on this test when handicapped populations are the subjects.

Picture Interest Inventory

The step-wise multiple regression analysis using the Geist Picture Interest Inventory yielded results that are similar to those using the VPI. The multiple correlation using the most predictive three sub-tests, Artistic, Clerical, and Literary yielded a multiple correlation of only .29, accounting for less than 10 per cent of the variance when the criterion was class rank. Thus, the VPI seems not to predict class rank assigned by teacher. Essentially similar results were obtained for prediction of ratings by teachers. The same caveats hold for these interpretations as are described for the VPI.

Purdue Pegboard

Step-wise multiple regression analysis of Right, Left, and Both Hands scores on the Purdue suggested that the prediction of class rank was trivial,

the multiple correlation being only .22, accounting for only about five per cent of the variance. It hardly seems worthwhile to spend the amount of time required to administer a test if one is to gain only a five per cent better than "best guess" chance of improving predictions. Again, it must be remembered that the criteria used were teacher assignment to class rank and teacher ratings and that these may or may not be predictive of later success in occupations. Nevertheless, in these results, only trivial predictions were found.

In an effort to determine the independence scores of the sub-scales of the Geist Picture Interest Inventory and Holland Vocational Preference Inventory, a factor analysis was used.

One would assume that if 11 sub-scales are used for a test, then 11 factors are being measured; interests measured by a test may be related, but the sub-scales should be sufficiently independent to provide bases for making decisions. If sub-scales overlap much, then some of them may be unnecessary.

For this population, the Vocational Preference Inventory reduced to three factors, although it is composed of 11 scales...six interest and five non-interest. The six "interest" sub-scales reduced to a single factor and the non-interest ones to two more factors. A single interest factor would seem to provide little usefulness for planning on the basis of interest because it seems to say only that there is "more or less interest" but not in what. In terms of this population, the Vocational Preference Inventory appears, then, to be either a one-factor test, not a good test for such populations, or an indication of some fantasy stage of development. Given the internal consistency procedure used by Holland in developing the Vocational Preference Inventory, the most probable conclusion is that it is not a suitable test for populations such as this one because on the basis of these results, one can say that if you know the interest sub-scale score on one sub-scale, you can predict the others.

The Geist Picture Interest Inventory reduced to five factors, the loadings on factors being what one would expect them to be; thus, as an interest measure, and based on factor analytical procedures, the Geist Picture Interest Inventory appears to be more appropriate for use with this population. However, there are qualifications that must be considered.

Ordinarily, one expects interest tests to produce results that are somewhat independent of intelligence; otherwise, one could simply base one's planning on IQ. The six Vocational Preference Inventory interest scales seem to be relatively independent of IQ, but the five non-interest scales are not. Strangely, the "intellectual" sub-scale of the Vocational Preference Inventory was not related to Full Scale Wechsler IQ! For several sub-scales of the Vocational Preference Inventory, there was an inverse relationship between the FSIQ and the Vocational Preference Inventory scale. One must question whether the "intellectual" sub-scale is indeed measuring "intelligence" as defined by Wechsler score. Furthermore, the Vocational Preference Inventory sub-scales for interest were trivial predictors of teacher ratings and of teacher assignment to class rank.

For the Geist Picture Interest Inventory, there were moderate positive correlations between sub-scales and Wechsler IQ, and these correlations were approximately what might be expected...scientific and computational sub-scales showed the highest correlations with IQ. Thus, again, for this population the Geist performance appeared to be more "reasonable than the Holland Vocational Preference Inventory one.

To investigate the relationships between sub-scales of the two tests combined, one can use factor analysis or a canonical regression analysis. Both were done. The factor analysis revealed little relationship between the Vocational Preference Inventory and the Geist Picture Interest Inventory sub-scales and those that existed came through the non-interest scales of the

Vocational Preference Inventory. The major relationship between the two measures is in measurement of mechanical, scientific, computational, and realistic interests. Structurally, the two tests appear to be measuring different things and are not interchangeable.

The canonical regression analysis assess the degree to which one set of variables is predictable (explainable) from another set of variables. Thus, it is a measure of the information redundancy of the two sets of information, or the degree of overlap and common information in the two sets of data. The canonical correlation between the 11 Vocational Preference Inventory and the 11 Picture Interest Inventory measures was about .79, suggesting that the two sets of data have about 50 per cent overlap when analyzed in this manner. This is a poor degree of overlap between the two instruments that are both (at least in part) supposed to be indicative of interests; if they were actually measuring similar or same interests, then one would expect correlations of the order of .90 to .95, and an 80 to 90 per cent overlap. Even the overlap that was found came through the "non-interest" sub-scales of the Vocational Preference Inventory, suggesting that what the Geist Picture Interest Inventory is measuring with its interest scales, the Vocational Preference Inventory is in part measuring with its "non-interest" scales. When a canonical analysis was run of the Geist Picture Interest Inventory and the first six (interest) sub-scales of the Vocational Preference Inventory, the multiple correlation dropped to .69, indicating an overlap of only about 47 per cent. Thus, one can say that at least for this population, using the canonical regression analysis, the finding was not different from the simple factor analysis, these two tests do not appear to have much in common.

Overview

In terms of all the data collected for this population, the structure among the variables appears to be as follows:

(1) The Wechsler IQ "explains" uniquely pieces of all the other variables. It was positively and fairly strongly related to class rank assignment by teachers, the average ratings by teachers, to handicapping condition, to prior grades in school, and to sex (males higher IQ than females). It was also positively related to the Purdue Pegboard performance. It was weakly but positively related to the two interest inventories.

(2) Handicapping condition was also a good predictor of class ranking by teachers and ratings by teachers. Thus, handicap and IQ, which is in part related to handicapping condition since some of these students were mentally retarded, were the best predictors of class rank and ratings by teachers of students.

(3) Prior academic achievement, as indicated by teacher assigned grades in the previous year, was a fairly strong predictor of class rank.

(4) Ratings by teachers was strongly related to class rank assigned by teachers. However, this may be in part a "halo" effect since the same teachers assigned both.

(5) The rest of the measures contributed relatively trivial information to the total picture, the interest inventories being not strongly related to any other variables. Whether they are related to later job performance cannot be predicted from these data. However, one might risk the speculation that failure to predict performance in a training class may also suggest something about later job performance, especially if the ratings on work performance by teachers can be assumed to be related to work performance by supervisors on the job at some later date.

In terms of predicting class rank assigned by teachers, the following seems, then, to make some unique contribution:

IQ

Sex

Type of Handicap

Prior grades on shop work

Ratings by teachers on work

The two single best predictors were IQ and handicapping condition; these were reciprocal, each having a correlation of about .60 with class rank assigned by instructors. Even so, much of the variance is still left unaccounted for; and so one must assume that other and perhaps more directly relevant measures are needed. Furthermore, one must consider the relationship between teacher assignment to class rank and later job performance. (A study is currently underway to attempt to determine the degree of relationship for students from one of the schools involved in this study, but that study is incomplete and not a part of Project VITA.)

VI. SUMMARY AND RECOMMENDATIONS

1. When consideration is given to the specific questions to be answered, one can improve efficiency by using selected standardized tests. Such questions might be those related to determination of skills (including intellectual and social skills) needed for jobs.

2. Selection of measurement devices must be directly related to the questions one has: Can the student learn the skill? Does the student have the motor coordination needed to perform the task? What interests does the student show that have implications for choice of training programs?

3. A very large number of published tests seem to have potential usefulness with one or more of the handicapped populations likely to be trained in vocational schools. A list of those that seem to have potential is provided and those planning to use tests may investigate the potential usefulness of any of these for determining information needed for helping students plan career and training choices.

4. The battery selected for use with the population of this project was chosen to provide the maximal amount of data with the minimal investment of time of students and professional personnel. Areas of relevant information were: Mental abilities, motor abilities, interests, selected personality variables, work and personal habits as rated by teachers. The major criterion against which these measures were measured was rank in class in present shop course.

5. The two most powerful predictors of class rank were IQ and handicapping condition.

6. The Purdue Pegboard correlated with IQ, but with none of the other variables measured and therefore it was considered as not providing any unique contribution in this battery with these students.

7. The Holland Vocational Preference Inventory was not highly correlated with other measures or with the criterion; factor analysis suggested that it probably is not measuring 11 variables, but only one interest and two non-interest variables. It did not appear to provide a unique contribution to the battery.

8. Teacher ratings on seven items were analyzed and two factors were found: Factor 1: Work Behaviors, and Factor 2: Personal Behaviors. Both of these were related to class rank, Factor 1 more so than Factor 2. Both factors can be hypothesized to be related to job success in the community, but this could not be determined in this project.

9. The Geist Picture Interest Inventory appears to be measuring relatively independent variables, as intended. It seems to have potential use with populations similar to this one.

10. Handicapping condition was a surprisingly good predictor of class rank. It was not possible to determine what factors went in to teacher determination of class rank assignment, but it seems unlikely that "prejudicial attitudes" could account for the consistency in which the finding occurred among the groups of teachers; therefore, it seems probable that the handicapping condition is a factor that needs to be given very serious consideration when helping handicapped students to plan vocational education. This is suggested by the fact that the rank order of class rank by groups was as follows: Speech/Hearing Handicapped highest; Learning Disabled second highest; Mentally Retarded third highest; and Emotionally Disturbed lowest. Generally, this order is one

that clinicians who have worked with the several types of handicapping conditions might agree seems to be one that is not unexpected.

REFERENCES*

Gardner, D. C. and Kurtz, M. A. Vocational Curriculum Models and Assessment Procedures Handbook for Handicapped Students. Final Report. Project VITA, Boston, Mass.: Boston University School of Education, Department of Business and Career Education, U.S. Office of Education, Project No. G007500558, Grant No. 498AH50296, June 1977.

Guilford, J. P. Fundamental Statistics in Psychology and Education (4th Edition), New York: McGraw-Hill Book Company, 1965.

Roscoe, J. T. Fundamental Research Statistics for the Behavioral Sciences, New York: Holt, Rinehart and Winston, Inc., 1969.

*All tests are referenced in Section III.

Boston University policy prohibits discrimination on the basis of race, sex, religion, age, color, creed, handicap, national and ethnic origin, and marital and parental status. This policy extends to all rights and privileges in the recruitment, admission, and employment of, as well as financial assistance to, students; the recruitment and employment of faculty and staff; and the operation of any of the University's programs and activities. Inquiries concerning the application of or compliance with Title IX and other state and federal laws and regulations prohibiting such discrimination should be addressed to the University's Director of Affirmative Action (881 Commonwealth Avenue, Boston, MA 02215) or to the Director of the United States Office for Civil Rights, or other appropriate government agency.